01

ERIM

FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN

II
7.7-1 0 1 6 9
CR-152670

WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA

TYPE II PROGRESS REPORT

16 November 1976 - 15 February 1977

114800-31-L

(E77-10169) WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA Progress Report, 16 Nov. 1976 - 15 Feb. 1977 (Environmental Research Inst. of Michigan) 41 p HC A03/MF A01 CSCL 02C G3/43

N77-23572

Unclas 00169

NASA Contract No. NAS5-22389

Prepared by

Richard F. Nalepka - Principal Investigator

John Colwell - Co-Principal Investigator

Daniel P. Rice

20622

Mr. G. R. Stonesifer, NASA Technical Officer/Code 902

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt Road
Greenbelt, Maryland 20771

WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA

TYPE II PROGRESS REPORT

16 November 1976 - 15 February 1977

The following report serves as the seventh Type II Progress Report for Landsat Follow-on Investigation #2062L which is entitled "Wheat Productivity Estimates Using Landsat Data".

This investigation has several objectives, including the following:

1. To develop techniques and procedures for using Landsat data to estimate characteristics of wheat canopies which are correlated with potential wheat grain yield.

2. To demonstrate the usefulness of Landsat data for estimation of winter wheat yield:

   a. for irrigated and for non-irrigated test sites

   b. for two different years with varying weather conditions

1.0 PROBLEMS

No significant problems were encountered during this reporting period.

2.0 SIGNIFICANT RESULTS

An initial demonstration was made of the capability to make direct production forecasts for winter wheat using early-season Landsat data. The approach offers the potential to make production forecasts quickly and simply, possibly avoiding some of the complexities of alternate procedures. Please refer to Section 8.

3.0 ACTIVITIES DURING THIS REPORTING PERIOD

During this reporting period, we investigated many of the important considerations that must be addressed in order to determine the usefulness of Landsat data for forecasting wheat productivity. In Section 4 of this report we discuss some of the fundamental relationships on which Landsat prediction of wheat yield are based, using measurements and observations of wheat condition which we made in the field during 1976. In Section 5 we discuss new Landsat data which was processed. We specifically address the relationship between Landsat data and yield, including optimal individual Landsat bands, optimal dates, and the relative usefulness of various Landsat green measure transforms. Section 6 addresses the issue of whether Landsat data is a useful indicator of yield relative to other sources of information. Section 7 is an examination of the extendability of relationships between Landsat data and yield over time and space. In Section 8 we describe and give an initial demonstration of a technique for direct estimation of total wheat production, which is an extension of our investigation of the relationship between Landsat data and yield.

4.0 FURTHER STUDY OF YIELD/LANDSAT DATA IN TERMS OF ERIM PERCENT COVER MEASUREMENTS

Field data reduction efforts continued during this reporting period, and all of the 1976 field photographs have now been reduced to measurements of vegetation cover. Before the results are presented, we will briefly review the procedure used to generate the data.

4.1 PERCENT COVER MEASUREMENTS

Initially, aerial oblique photos were taken of selected fields. Fields to be sampled on the ground were then selected from the aerial photos so as to furnish a range of field conditions, vegetative cover, and probable yield. The fields were internally stratified using the aerial obliques so that samples within each field would represent the range of conditions in the field.

Vertical photographs of the fields were then obtained on the ground. Up to eight photos per field were obtained, depending on the variability within the field. The photographs were projected onto a large screen, and the proportion of the canopy representing the following components was determined:

1. green leaves
2. green stalks
3. green heads
4. green weeds
5. senescent leaves
6. senescent stalks
7. senescent heads

The above categories were aggregated into various combinations. The most commonly discussed combination is green wheat cover, which is composed of Items 1-3.

Once the individual photographs were reduced to the seven categories, the data was used to produce estimates of vegetation condition for the entire field. The field was divided into relatively homogeneous areas (strata), each represented by the measurements from one or more of the photographs. The proportion of the field occupied by each stratum was determined. Then the individual stratum average values were multiplied by the corresponding stratum proportion and aggregated to produce a single value characteristic of the field. In the following discussion, percent cover measurements will refer to measurements of percent green wheat cover, unless otherwise stated.

4.2 LANDSAT/PERCENT COVER RELATIONSHIPS

Since our Landsat yield prediction methodology is based on Landsat data being a good indicator of green development, we examined the relation between a Landsat green feature indicator and the ERIM field measurements of percent green wheat cover.

The relationship between condition (percent green wheat cover) and Landsat data was investigated for the two dates for which both Landsat data and field data were available, namely 18 April 1976 and 2 June 1976. On 18 April essentially all of the vegetation was green, whereas on 2 June many fields contained appreciable amounts of senescent vegetation (up to 30% dead wheat cover).

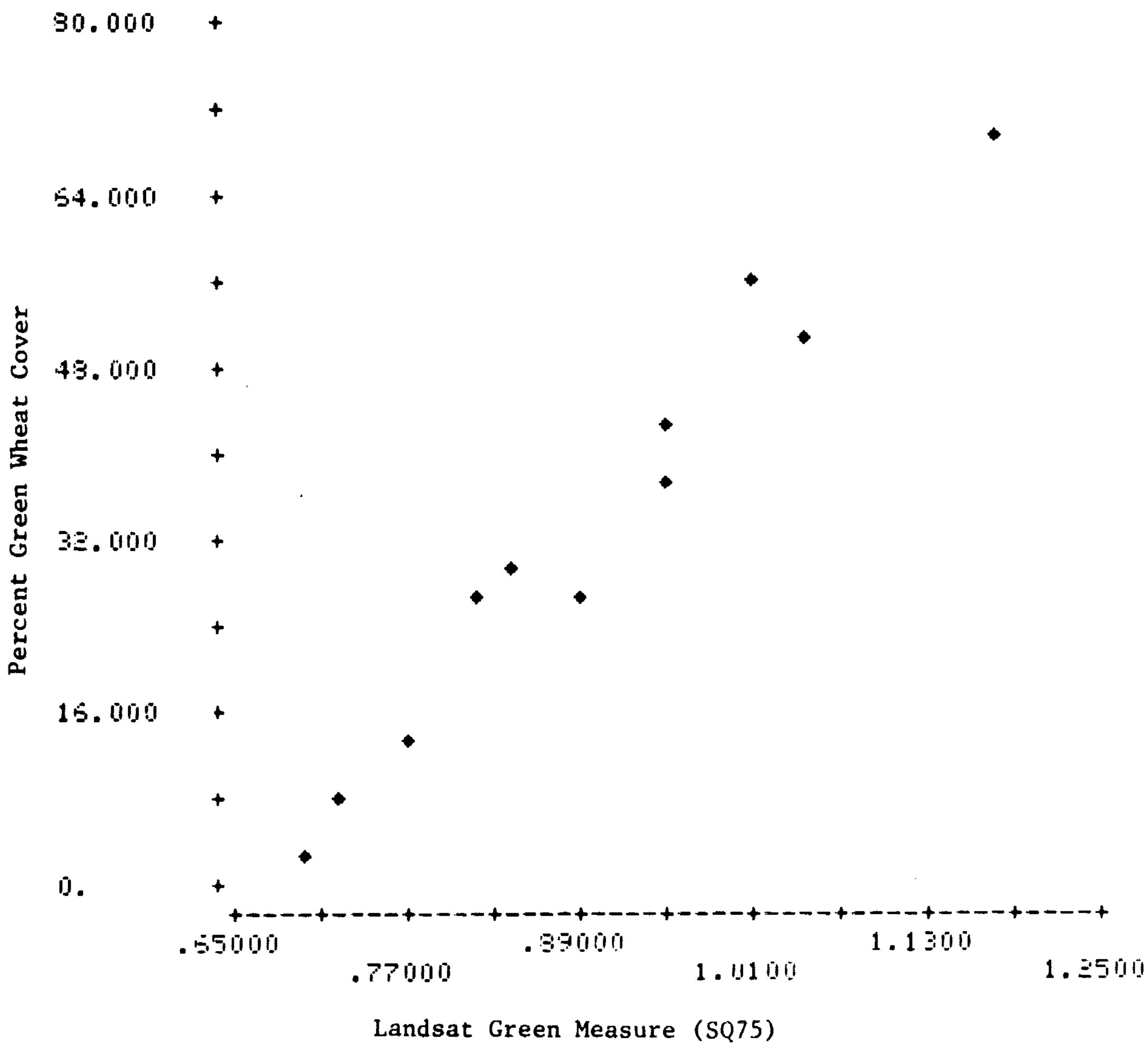
For the 18 April data the correlation between the square root of an MSS7/MSS5 ratio (called SQ75) and ERIM estimates of percent green wheat cover for 11 fields was 0.98. This is highly statistically significant (see Figure 1). For the 2 June data the correlation between SQ75 for 12 fields was 0.79, somewhat lower than for 18 April, but still highly significant in the statistical sense. Because farmers plowed some fields between 18 April and 2 June, only eight of the sampled fields were the same for the two dates. For these eight fields the correlations with SQ75 and percent green wheat cover were 0.97 for 18 April data and 0.78 for 2 June data. Both correlations are significant, but for 2 June data this is so only at the 5% level.

While a more definitive conclusion would await evaluation of a more extensive data set, the 18 April 1976 Finney results increase our confidence that Landsat data can provide a good indication of the amount of green vegetation cover when the wheat is predominantly green. A tentative conclusion to this effect was made on the basis of 21 May 1975 Finney data in a previous quarterly report. In addition, the 2 June results suggest that a Landsat green indicator may work reasonably well for prediction of percent cover in partially senescent wheat canopies, although probably not as well as in all-green canopies.

4.3 TEMPORAL ASPECTS OF PERCENT COVER AND WHEAT PHENOLOGY

We hypothesized that the optimum single date for forecasting probable yield by estimating field condition (vegetation cover) using Landsat data is approximately at the time of heading. It was assumed that heading date corresponded approximately with the time of maximum vegetation cover. This section examines the timing of heading and vegetative development between and within fields which we examined for the 1976 Finney site.

FIGURE 1. SCATTER PLOT OF A LANDSAT GREEN MEASURE (SQ75) vs ERIM FIELD MEASUREMENTS OF PERCENT GREEN WHEAT COVER (18 April 1976 Finney Site)



The timing of heading and vegetative development was extremely variable at the Finney County ITS in 1976. An indication of the temporal variability in relative field condition is that for the eight fields examined on both 18 April and 2 June, the correlation between the measurements of percent green wheat cover is only 0.06. Clearly the relative condition of the individual fields has changed considerably during this period of time. This variability in relative field condition as a function of time suggests that multiple looks at crop condition may be important for these fields for accurately forecasting yield.

Our field observations indicated that some of the fields were almost completely headed on 14 May, whereas other fields were not completely headed by 2 June. In addition, some fields reached peak vegetative cover before 14 May, and did not head until considerably later. Our field measurements of percent green wheat cover indicated that four of the fields sampled on both 18 April and 14 May had less green wheat cover on 14 May than on 18 April, whereas four other fields had greater green wheat cover on 14 May. Furthermore, there were even variations in timing of heading and peak vegetative cover within a given field. For example, in one field the dense portions of the field decreased from 62% green wheat cover on 18 April to 41% vegetative cover on 14 May, while on the sparse portions of the same field the green wheat cover increased from 29% on 18 April to 36% on 14 May.

The considerable variability in phenology for the fields which were observed, even though meteorological conditions for all fields were probably quite similar, suggests that being able to accurately account for variations in phenology based on meteorological factors (e.g., day and night temperatures and photoperiod) may not always be possible.

4.4 PERCENT COVER/YIELD RELATIONSHIPS

As a result of the complex pattern in the relationship of percent cover and heading date between and within the fields for which we made detailed field observations, we anticipated an uncertain relationship between percent cover and yield, and hence between Landsat data and yield for those fields. This uncertain relationship was confirmed by further analysis.

The ERIM field measurements of percent green wheat cover were correlated with yield for all three dates for which data was available, namely 18 April, 14 May, and 2 June. None of the dates showed statistically significant correlations between percent green wheat cover and yield. Similarly, the correlation of yield and SQ75 is not significant for 18 April, 6 May, and 12 June, and only barely significant for 2 June.

It should be noted that the relationship between Landsat data and percent cover is a much more straightforward relationship than the relationship between percent cover and yield or Landsat data and yield. The relationship between Landsat data and percent cover is basically a physical-electromagnetic relationship which connects two observations at the same point in time in a more or less causal fashion [1]. Yield, on the other hand, has a much more complicated relationship with field condition which has a strong plant physiological component and which is the integrated effect of a host of conditions over time. The relationship may be affected by conditions occurring before or after the observation of field condition. For example, one of the fields on which we made measurements developed significant mosaic virus which undoubtedly altered the relationship between vegetation cover at a point in time and yield. In addition, as noted previously, observations at one point in time do not necessarily compare fields at similar phenological stages.

The significant point is this: an accurate measure of percent green wheat cover (or stand quality) at a point in time (whether from field measurements or Landsat data) does not guarantee a good measure of eventual wheat grain yield on all fields at all times. The potential success of such an approach depends on the relationship being generally useful most of the time. It also depends on being able to correct yield estimates at a point in time for future anomalous conditions (e.g., disease, hail) if they should occur. The same limitation is imposed on any system or approach in which the future is uncertain.

As will be shown in Section 5.2, for the large population of fields for which we have both Landsat data and yield, there is a significant correlation between Landsat green transforms (such as SQ75) and yield for all dates. This situation suggests that the small population of fields for which we made field measurements is somewhat anomalous with respect to the relationship between field condition and yield. Therefore, on this site, despite certain anomalies, Landsat indicators of field condition (and presumably field measurements) are generally useful indicators of yield.

4.5 GREEN COVER DURATION

One of the hypotheses we examined previously is that percent green wheat cover integrated over time is more highly correlated with yield than data on a single "optimum" date. Previous results failed to show this to be true [2]. We address the question again using newly prepared 1976 Finney data. Since results reported in Section 4.4 suggest that the fields on which we made ground measurements and observations were somewhat anomalous with respect to yield compared to the site as a whole, we chose to analyze only the larger set of fields using a Landsat green indicator as a surrogate for percent green wheat cover.

The Landsat green indicators from 18 April, 6 May and 2 June were summed to approximate percent cover over that span of time. This sum was then correlated with wheat yield, and the correlation was found to be 0.89. The best single date has a correlation between Landsat data and yield of 0.81. The 12 June data was not used in this example because there is very little green vegetation present at this time of year, and because Landsat is not expected to be a good green indicator under such conditions. If 18 April, 6 May, and 2 June are used as independent variables for regression with yield, the multiple correlation is 0.92, somewhat better than if the data were summed.

In summary, our results for both the 1975 data (reported previously) and the 1976 data indicate that useful information can be achieved by using more than one date of Landsat data. However, it does not appear as though a summation or integral of Landsat green indicator over time is the optimum way to use the information in multiple Landsat passes.

5.0 FURTHER STUDY OF THE RELATIONSHIP BETWEEN YIELD AND LANDSAT DATA

During this reporting period additional Landsat data has been prepared for analysis of its relationship to wheat grain yield. Details of this analysis is presented in this section.

5.1 ANALYSIS OF OPTIMAL LANDSAT BANDS

With the addition of three new Landsat data sets, sixteen Landsat spectral-temporal bands were available for the 1975-1976 Finney County site. The four Landsat data acquisitions which were analyzed were 18 April, 6 May, 2 June, and 12 June 1976.

As before, mean signal values in each Landsat band were computed for each sufficiently large wheat field, and these values were subsequently correlated with the farmer estimates of wheat grain yield (per planted acre) in order to assess relative information content.

The results for the cases for which suitable Landsat data and yield was available are presented in Figure 2. For all four dates, both visible bands (MSS4, MSS5) were significantly negatively correlated with yield. MSS7 was significantly positively correlated with yield for all dates except 12 June, the date when much of the canopy was senescent. On 12 June, MSS6 was significantly negatively correlated with yield.

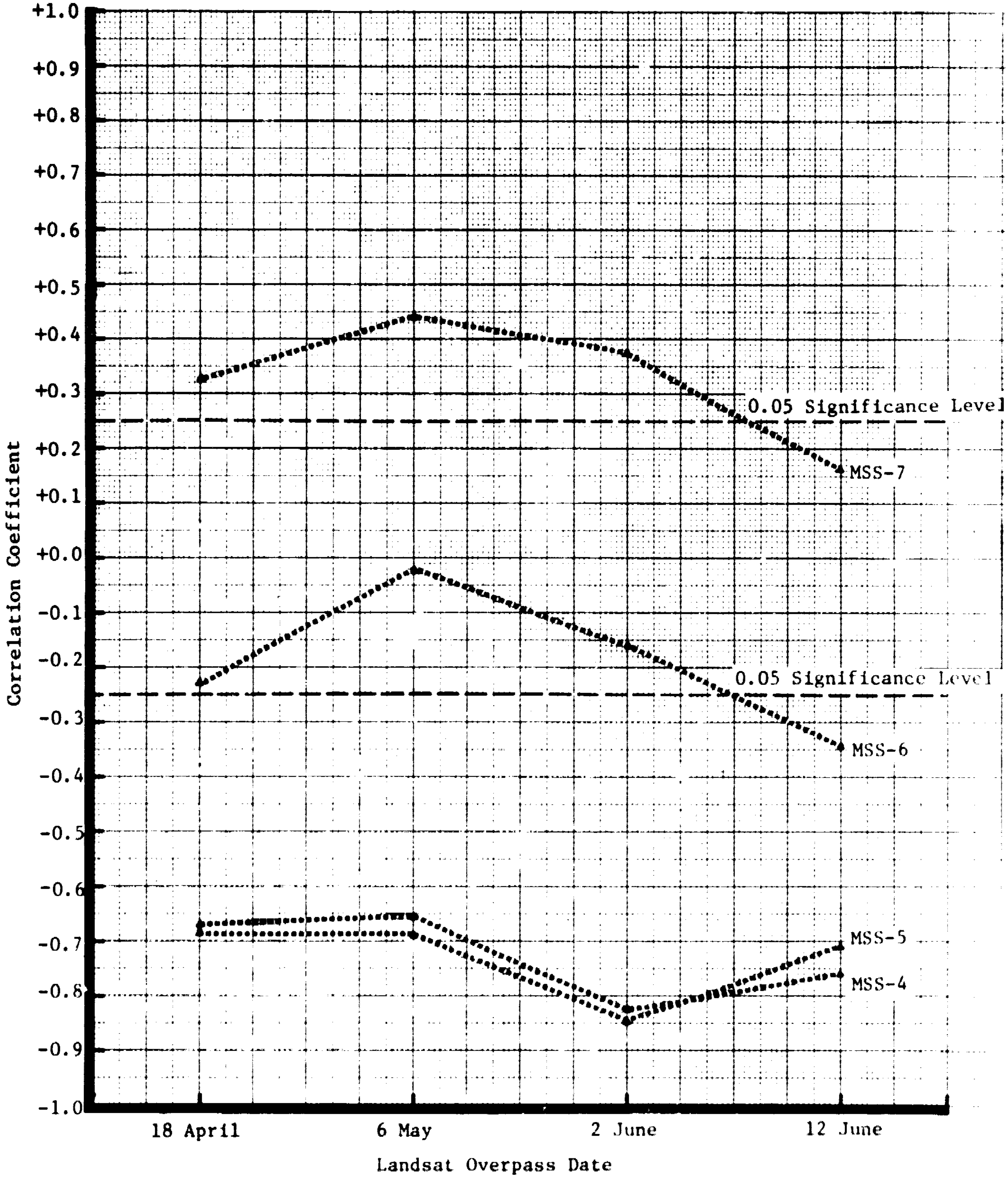
Of the two IR bands, MSS7 is clearly a better indicator of yield than MSS6 for the 1976 Finney data. There is little difference between the two visible bands, although MSS5 (red band) tends t[illegible] slightly superior to MSS4 (green band). These results are similar to those found on previous test sites.

Using these same data, the optimum spectral-temporal bands for predicting yield for 1976 Finney data were determined by stepwise regression. The result of the regression indicated that the four optimum spectral-temporal bands* came from the 18 April and 2 June

---

*18 April Bands 5,6; 2 June Bands 5,7.

FIGURE 2. CORRELATION OF INDIVIDUAL LANDSAT BAND DIGITAL COUNT VALUES vs YIELD AS A FUNCTION OF DATE. 1976 FINNEY SITE



acquisitions. These four optimum bands accounted for more than 83% of the variance in yield as measured by the coefficient of determination (R^2).

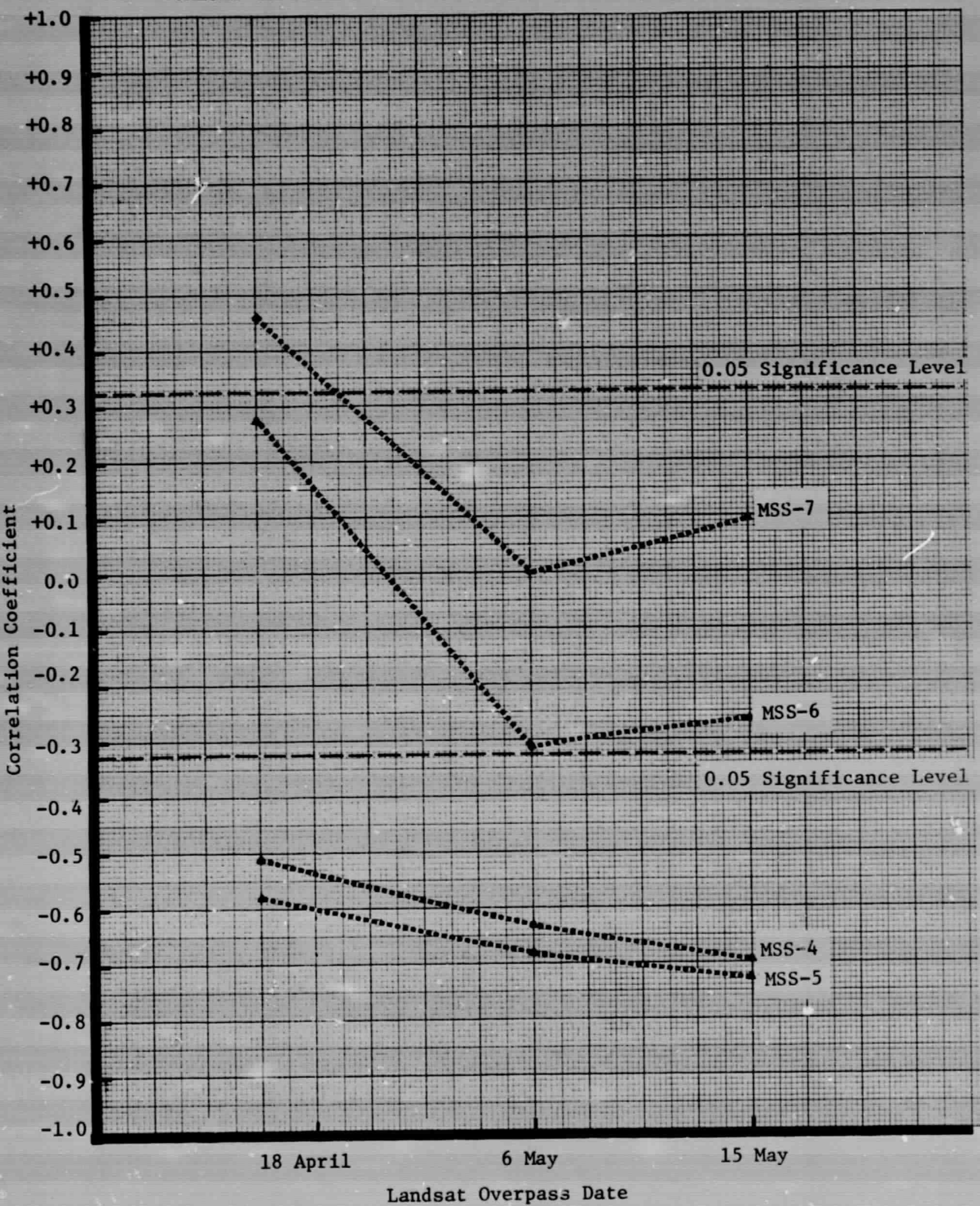
In order to determine the best single date for predicting wheat grain yield using all four bands on a given date, a regression was formed between yield and the set of four bands for each date. These regressions indicate that 2 June is the optimum date of those investigated. The four Landsat bands from 2 June account for 74% of the variance in yield, with a standard error of estimate of 6.5 bushels per acre.

However, 18 April, 6 May and 12 June data are all nearly the equivalent of 2 June, having coefficients of determination (R^2) of 0.66, 0.64 and 0.72, respectively. It is clear that several dates of Landsat data furnish important information related to yield. The importance of the late season (2 June, 12 June) data is likely due to the importance of late season crop development (after heading), as was indicated by ERIM field observations (see Section 4.3).

We expect heading to be the optimal date for correlation between Landsat data and yield. Since suitable 15 May Landsat data was not available, it is not possible to assess the validity of our expectation for the 1976 Finney site.

In addition to the above data, three acquisitions of Landsat data have been processed for the 1975-76 Ellis site, namely 18 April, 6 May and 15 May. The correlation between individual Landsat band field mean values and yield is shown in Figure 3. In this case, the visible bands (MSS4, MSS5) are significantly correlated with yield for all three dates, and MSS5 is somewhat more highly correlated with yield than MSS4 on each of the dates. Landsat MSS7 is significantly correlated with yield only on 18 April, and MSS6 is not significantly correlated with yield on any of the dates. This behavior on the part of the IR bands is contrary to our expectations. We would expect fields

FIGURE 3. CORRELATION OF INDIVIDUAL LANDSAT BAND DIGITAL COUNT VALUES vs YIELD AS A FUNCTION OF DATE. 1976 ELLIS SITE



with high yield to have significantly greater amounts of green wheat cover and correspondingly larger IR radiances (especially MSS7) than fields with low yield. We have not yet determined the reason the observations differ from what was expected.

The optimum spectral-temporal bands for predicting yield for the 1976 Ellis site were determined by stepwise regression, as before. The resulting four best spectral-temporal bands were from 18 April and 15 May*, and accounted for 60% of the variance in yield.

The best single date for predicting wheat grain yield was determined by regression between yield and the set of four bands for each date. These regressions indicate that all 3 dates were approximately equivalent, but that 6 May was slightly superior to the other two. Early to mid-May is approximately the time at which heading occurred, and that is when we would expect maximum correlation between Landsat data and yield. Earlier and later Landsat data will have to be examined in order to assess whether our expectations of optimum date are confirmed for this site.

5.2 ANALYSIS OF TRANSFORMS TO EXTRACT GREEN COVER INFORMATION

The results of a regression of Landsat individual bands with yield on one site indicates an upper limit of yield-prediction performance that could be achieved if the relation were applied to another site on which all conditions were the same. However, conditions are never quite the same on another site, and yield-prediction performance is normally somewhat degraded.

We are investigating methods of minimizing variability in Landsat signals due to such things as variable soil reflectance and atmospheric scattering. As discussed in previous quarterly reports, one of the ways we have approached this problem is by implementing green feature transforms which tend to accentuate differences in green vegetation

---

*18 April Bands 5,6; 15 May Bands 5,7.

cover and minimize other differences. As will be shown later, transforming data to minimize these differences tends to result in some reduction of yield-predictive capability locally.

Various green measure transforms (See Table 1) were implemented for the four 1976 Finney Landsat dates and for the three 1976 Ellis Landsat dates. The transforms were correlated with farmers' yields for each of the dates. For 1976 Finney data, SQ75 was generally slightly superior to other transforms tested, and for 1976 Ellis data TVI was slightly superior. The results of the respective optimal transforms for the two sites are presented in Figures 4 and 5. Note that the apparent "optimal dates" for the transforms are the same as was found using untransformed Landsat data. In addition, the data indicate that important yield-predictive information is present at several points in time including post-heading dates.

TABLE 1. DEFINITION OF GREEN FEATURE TRANSFORMATIONS

Transformation Name	Definition
R75	$\frac{MSS7}{MSS5}$
SQ75	$\sqrt{\frac{MSS7}{MSS5}}$
TVI	$\sqrt{\left(\frac{MSS7 - MSS5}{MSS7 + MSS5} + 0.5\right)}$
G*	MSS4 - MSS7 + 96

*Discussed in Section 5.4

FIGURE 4. CORRELATION OF GREEN MEASURE (SQ75) VALUES vs YIELD AS A FUNCTION OF DATE. 1976 FINNEY SITE.

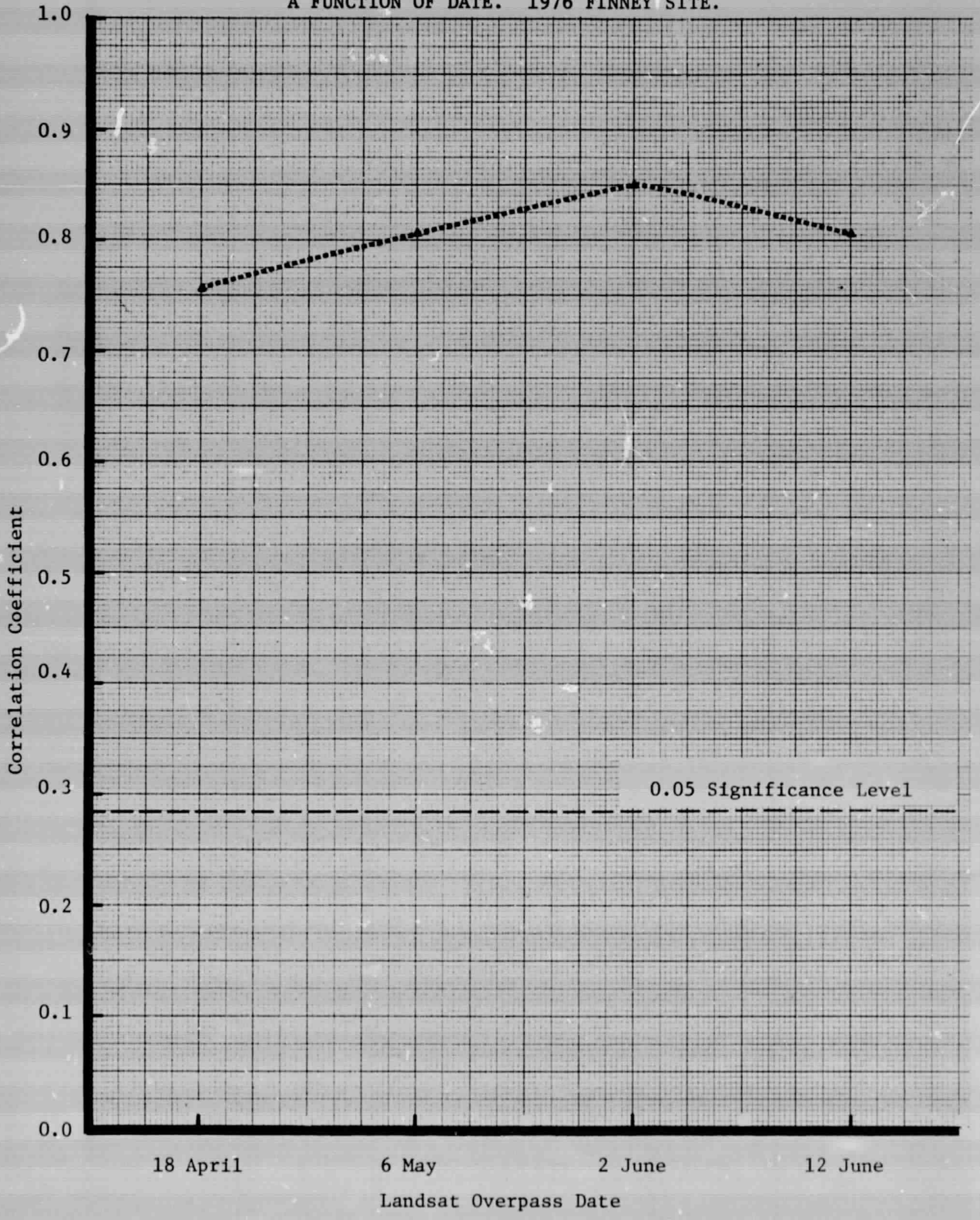
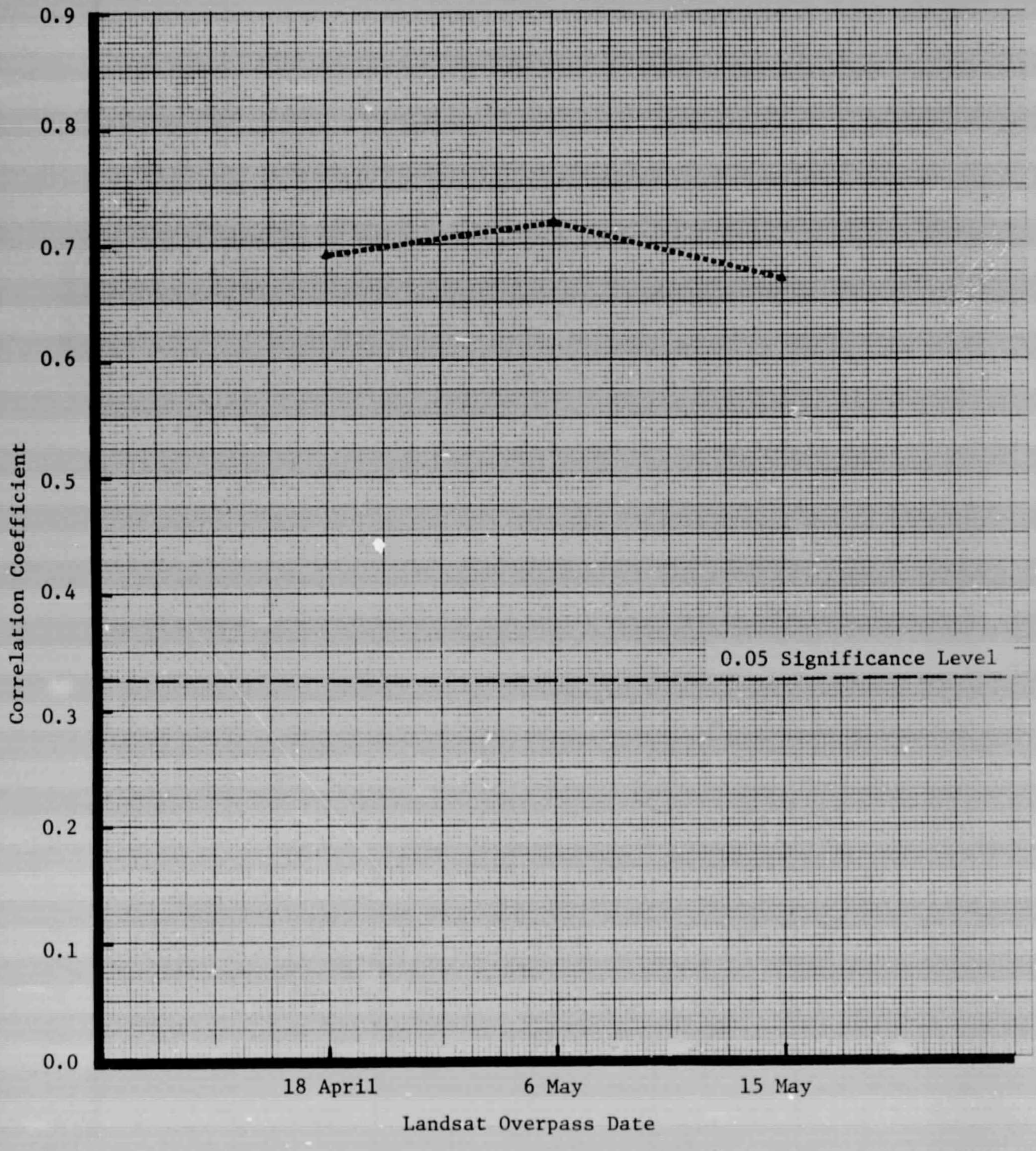


FIGURE 5. CORRELATION OF GREEN MEASURE (TVI) VALUES vs YIELD AS A FUNCTION OF DATE. 1976 ELLIS SITE.



5.3 ANALYSIS OF LATE-SEASON LANDSAT INDICATORS OF YIELD

We cannot a priori expect a green indicator transform to be highly correlated with wheat yield when the wheat fields have essentially no green wheat present, as is the case in senescent canopies. Thus far, we have processed two sets of Landsat data in which the wheat was largely senescent, namely 12 June 1976 Finney and 17 June 1975 Ellis. A comparison of individual band and green indicator transforms for the two data sets is presented in Figure 6.

In both cases the green indicator transform is significantly correlated with yield, though barely so on the 17 June 1975 Ellis data. However, in both cases 3 of the 4 bands (MSS4, 5, 6) are significantly negatively correlated with yield. Furthermore MSS4, 5 and 6 are all more highly correlated with yield individually than is the green indicator transform. It appears as though wheat yield is negatively correlated with crop albedo when the crop is mature. Perhaps this is due to high-yield fields having more stalks and hence casting more shadow (having lower reflectance) than low-yield fields.

The reason that a green indicator transform is still significantly correlated with yield in such a situation is not clear, and must be further studied. However, it appears that for late-season (pre-harvest) estimates of yield, some albedo estimator or other stand density estimator could be a better indicator of yield than a green indicator.

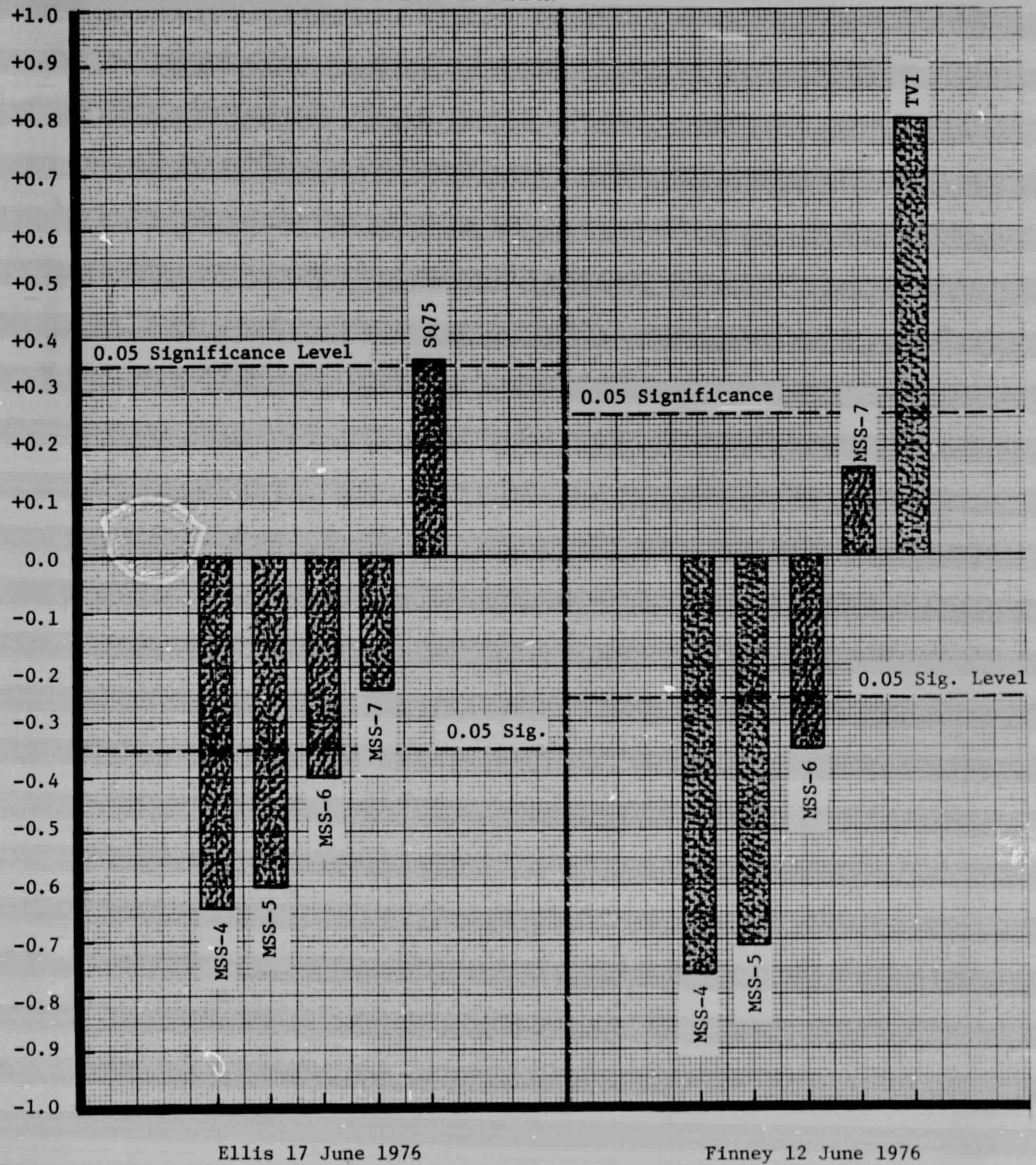
5.4 GREEN MEASURE G

Some investigators have indicated success in detecting wheat using the following transformation of Landsat data:

$$G = (\text{Landsat Band } 4) - (\text{Landsat Band } 7) + 96$$

It has been suggested that this transformation (which we call "G") may be a good measure of green vegetative cover, and that it may be relatively little affected by external conditions such as haze. Because of the importance associated with these factors in terms of the present

FIGURE 6. CORRELATION OF INDIVIDUAL BAND AND GREEN MEASURE VALUES vs YIELD LATE IN SEASON



investigation, a joint effort was begun between this contract and another* to study the utility of the G transformation.

The first test was to study the extent to which wheat is separable from non-wheat using the G transformation. For 6 May 1976 Landsat data, G was computed for 96 fields. Histograms produced for the wheat fields and the non-wheat fields are compared in Figure 7. While there is some overlap in the two distributions, the transformation has largely separated the two classes, as expected. We then performed the same test using two of the green measures we have studied previously during this investigation (TVI and SQ75). The result is that little difference in the separability of wheat and non-wheat using the various transforms was found. Thus at least for purposes of recognizing wheat, it appears as though these transformations including G, are roughly equivalent.

The G transformation was subsequently analyzed as a quantitative measure of green cover and yield. For 21 May Landsat data on the 1975 Finney site the transformation was highly correlated with both percent green cover and leaf area index, but not as significantly as some of the other green feature indicators we have investigated. It was also highly correlated with yield, but again not to the same degree as other green feature indicators. The same situation was found to be true for both 20 May and 21 May 1975 Ellis data, and also for 6 May 1976 Finney data.

To further study the relationship between G and green vegetative cover, and also to test the sensitivity to external effects, we computed the transformation on simulated Landsat data which was generated using the ERIM Canopy/Atmospheric Model [3]. In addition, the transformations TVI and SQ75 were similarly computed on the simulated data. Nine separate canopies were modeled, each having its own value of percent cover.

---

*Contract #NAS9-14988 with NASA/JSC.

FIGURE 7. HISTOGRAMS OF G TRANSFORM SHOWING SEPARABILITY OF WHEAT AND NON-WHEAT CLASSES. FINNEY SITE, 6 MAY 1976.
(Each x = 1 Field)

G Value	Non-Wheat	Wheat
82.000	0 +	0 +
83.000	0 +	3 +XXX
84.000	0 +	0 +
85.000	0 +	2 +XX
86.000	0 +	2 +XX
87.000	0 +	3 +XXX
88.000	0 +	3 +XXX
89.000	0 +	1 +X
90.000	0 +	4 +XXXX
91.000	0 +	5 +XXXXX
92.000	0 +	4 +XXXX
93.000	0 +	6 +XXXXXX
94.000	0 +	4 +XXXX
95.000	0 +	7 +XXXXXXX
96.000	0 +	1 +X
97.000	0 +	1 +X
98.000	1 +X	2 +XX
99.000	0 +	2 +XX
100.00	0 +	2 +XX
101.00	0 +	1 +X
102.00	0 +	1 +X
103.00	4 +XXXX	0 +
104.00	2 +XX	2 +XX
105.00	3 +XXX	0 +
106.00	6 +XXXXXX	1 +X
107.00	3 +XXX	1 +X
108.00	9 +XXXXXXXXX	0 +
109.00	4 +XXXX	0 +
110.00	5 +XXXXX	0 +
111.00	1 +X	0 +
112.00	0 +	0 +

When external factors were held fixed, we found that the correlation between percent cover and G using modelled Landsat data was 0.97, and that the standard error in estimating percent cover using G was 8.1 percentage points. The comparison shown in Table 2 indicates that the G is roughly comparable, but slightly superior, to two other transformations for measuring percent cover.

TABLE 2. COMPARISON OF G AND OTHER TRANSFORMATIONS FOR MEASURING PERCENT COVER USING MODELLED DATA (9 Points)

Green Measure	Correlation With Percent Cover	Standard Error in Measuring Percent Cover
G	0.97	8.1
TVI	0.91	12.8
SQ75	0.95	9.6

Again using model-simulated Landsat data, we examined the variation in G one should expect due to normal variations in haze, view angle, and background albedo. For each canopy, a Landsat signal was computed for each of several conditions of each of the four external parameters under consideration, resulting in a total of about 1200 points total. Using these points, a regression was run relating percent cover and G transform value. The result was a standard error of 23.7 percentage points in estimating percent cover. Using the same procedure with the transforms TVI and SQ75, the corresponding standard errors were 19.9 and 20.8.

In summary, we found that the G transform is largely comparable to other transforms in terms of measuring green cover and potential yield and in terms of sensitivity to variations in external conditions.

While G tended to be slightly more sensitive to green cover, it simultaneously tended to be slightly less resistant to external effects.

6.0 RELATIVE UTILITY OF LANDSAT AND ALTERNATIVE SOURCES OF INFORMATION FOR ESTIMATING WHEAT YIELD

In the previous sections we have examined the utility of Landsat data for estimating yield. The value of using Landsat data in an actual wheat survey can to some extent be judged by comparison with the use of other sources of information. In addition to this comparison, Section 6 addresses the possibility of using a combination approach, in which more than one type of information would be simultaneously used for assessing yield.

6.1 LANDSAT DATA VERSUS ALTERNATIVE INFORMATION

We have noted in a previous quarterly report [2] that some yield models require as inputs certain estimates of vegetation condition. Therefore, we will compare several sources of data (including Landsat) in terms of ability to measure percent cover. As a basis for comparing percent cover estimates, the carefully made ERIM objective field measurements, described in Section 6.1, will be used. For the 1976 Finney site, the 18 April ASCS subjective estimates of percent cover and the ERIM field measurements of percent cover have a correlation of 0.71. The corresponding correlation between Landsat data (SQ75) and ERIM measurements is 0.97. This result adds some credence to our previous preliminary conclusion that for yield models that require estimates of degree of crop vegetative development, Landsat data may furnish a better estimate than some subjective estimates made by field personnel using traditional approaches.*

*Traditional methods using trained field personnel can certainly be made more precise than Landsat data, but the traditional methods are sufficiently time-consuming so that they cannot routinely be made on enough samples to characterize large, variable fields.

The correlations between various estimates of field vegetative condition and actual yield are shown in Table 3. None of the correlations with yield are statistically significant for this sample. However, the correlations are highest for ERIM objective measurements of green cover and for Landsat data (SQ75). This result is similar to one achieved using 1975 Finney data [4].

TABLE 3. CORRELATIONS BETWEEN VARIOUS INDICATORS OF CROP CONDITION AND YIELD, 18 APRIL 1976, FINNEY DATA (N = 9)

Variable	Correlation With Yield
Percent Cover (ASCS)	0.18
Height (ASCS)	-0.17
Green Cover (ERIM)	0.52
SQ75	0.45

Significance (0.05) = 0.67

In the previous quarterly report [2] we investigated the relative correlation with yield for Landsat data and for alternative traditional means of estimating yield. With the addition of more information on yield and stand quality ratings (SQR) obtained during this reporting period, a more complete analysis of 1976 Finney data was made. The results are presented in Table 4.

TABLE 4. CORRELATIONS WITH YIELD OF INDIVIDUAL FIELDS FOR LANDSAT DATA AND TRADITIONAL ESTIMATES, FINNEY 1976 SITE, 55 FIELDS

Estimate	Date	Correlation
FCIC Yield	Pre-harvest	0.26
Stand Quality Rating (SQR)	Pre-harvest	0.78
Landsat Data (SQ75)	6 May	0.82

These results, based on a larger data set than used previously, strengthen our previous preliminary conclusion that Landsat estimates of probable yield are as well correlated with actual yield as are some traditional in-the-field alternatives, even for Landsat data collected well before the estimates using alternative methods.

We now examine the wheat yield information accounted for by cultural factors on the 1976 Finney site, and the degree to which Landsat data monitors their effects.

The cultural practices investigated included:

1. wheat variety
2. irrigation (yes/no)
3. fertilization (yes/no)
4. planting date
5. summer fallow (yes/no)
6. amount of fertilizer (lbs per acre)

All of these variables are potentially available early in the growing season, and hence could be available for early yield forecasting.

An analysis of variance was performed for the above factors by linear regression with wheat yield for the 55 fields for which such data was available. From this analysis, it was possible to determine the percent of variance in yield accounted for separately by each of the factors. However, high correlations do exist between some of the variables, so the results cannot be treated as though the variables were independent of each other. The results are presented in Table 5.

Planting date, somewhat surprisingly, accounts for almost none of the variance in yield on these particular fields. Perhaps the over-wintering period tends to reduce potential differences due to planting date.

Wheat variety accounts for only a small amount of yield variance. This is to be expected, because the principle wheat varieties planted on this site (Eagle, Scout, and Centurk) have similar "yielding abilities" [5].

TABLE 5. PERCENT OF VARIANCE IN YIELD ACCOUNTED FOR SEPARATELY BY SEVERAL CULTURAL FACTORS, 1976 FINNEY SITE

Cultural Factor	Percent of Variance
Planting Date	0.1
Wheat Variety	10.6
Previous Cropping	35.8
Irrigation	56.3
Fertilization	55.0
Amount Fertilization	57.4

Previous cropping practice (whether the field was summer fallowed) accounts for an appreciable amount of variance in individual field yield. This is not unexpected since the reason for leaving a field fallow is to improve the soil characteristics for the subsequent crop.

Irrigation, fertilization, and amount of fertilization, all account for a substantial amount of variance in yield. They are highly correlated with each other, however, and the three variables combined do not account for much more variance than each one individually.

The amount of variance accounted for by a Landsat green indicator (SQ75) for each of the four dates processed was computed for the same fields that were used in the above analysis. The results are presented in Table 6. Landsat data from either 6 May, 2 June, or 12 June account for more variance in yield than any single cultural factor examined.

TABLE 6. PERCENT VARIANCE ACCOUNTED FOR SEPARATELY BY SEVERAL DATES USING SQ75 (1976 FINNEY ITS)

Date	Percent of Variance
18 April	54.8
6 May	67.7
2 June	72.0
12 June	67.4

6.2 COMBINATIONS OF DATA FOR PREDICTING YIELD

In Section 6.1, we discussed the usefulness of various data sources for predicting yield. In this section, we address the question of predicting yield using data from selected combinations of sources.

Table 7 gives the results. Note that, together, all of the cultural variables (1-6) account for a substantial amount of yield variance (75%). Nevertheless, the Landsat green indicators for the four dates (7-10) account for even more variance in individual field yield (87%) than all of the cultural variables. The combination of all Landsat and cultural variables accounts for almost all of the variance in yield (94%).

TABLE 7. PERCENT OF VARIANCE IN YIELD ACCOUNTED FOR BY SEVERAL COMBINATIONS OF CULTURAL AND LANDSAT VARIABLES

Variables	Percent Variance	Standard Error
1-6 (all cultural vars)	74.9	6.89
7-10 (all Landsat vars)	87.3	4.78
4,5,7,10 (optimum four vars)	90.7	4.10
1-10 (all vars)	93.6	3.65

Variable Key

1 = variety	6 = amount fertilizer
2 = irrigation	7 = SQR75 (May 6)
3 = fertilization	8 = SQR75 (June 2)
4 = planting date	9 = SQR75 (June 12)
5 = cropping	10 = SQR75 (April 18)

We had previously speculated that field condition as measured by Landsat would account for the integrated effects of the factors governing crop growth and potential yield, including the cultural factors. The effect of cultural factors are most clearly seen on a local area where meteorological conditions are similar, and these cultural factors are almost completely accounted for by Landsat data in this 1976 Finney site. For example, addition of all six cultural factors to the four Landsat variables increased the variance accounted for by only 6.3%.

The standard errors of estimate are also worth noting. Using the four Landsat variables the standard error is 4.78 bu/acre on this test site. If this performance could be achieved on 100 randomly selected fields with a normal distribution of yields about the mean, the average yield on the 100 fields could thus be estimated to within $\pm$0.478 bu/acre, a significant potential accomplishment.

6.3 METEOROLOGICAL DATA

As discussed in the previous quarterly report, we believe that there is important information for forecasting wheat yield contained in Landsat data that is not provided by standard meteorological data. Meteorological conditions were undoubtedly relatively similar over the 5 x 6 mile 1976 Finney site [2]. And in fact, aside from special experimental arrangements, a single weather station covers an area much larger than a single test site, and thus would not indicate any field-to-field differences in yield based on meteorological conditions. The yield, however, varied substantially from field to field on this site (from 3 bu/acre to 65 bu/acre).

While we do not down-play the usefulness of using meteorological information to roughly estimate yield on a regional average basis, or to help assess approximate status of phenological development, we feel that accuracy of a large area wheat survey could be enhanced by the use of field-by-field information, such as could be provided by Landsat data.

7.0 YIELD PREDICTION EXTENSION

We have continued investigation of the feasibility of extending the Landsat data - wheat yield relationship developed under one set of conditions (environmental, cultural) to Landsat data collected under different conditions at a different place and/or time. In this section of this quarterly report we discuss four such tests of 1976 data:

1. 18 April Ellis to 18 April Finney
2. 6 May Ellis to 6 May Finney
3. 18 April Finney to 18 April Ellis
4. 6 May Finney to 6 May Ellis

Extensions were made from Ellis to Finney and from Finney to Ellis for both dates, because the direction of the attempted extension sometimes affects results.

Extensions are carried out by establishing a relation between Landsat data (or a transformation) and yield on one site, and applying the relation on another. This was done using

1. All four Landsat bands
2. SQ75
3. TVI.

The latter two quantities are transformations designed to measure green cover, and were described in Section 5.2. They were chosen rather than other transformations we have tested because they generally have been found to perform slightly better.

In order to implement Landsat/wheat yield relations on analogous stages of phenological development on the two sites, we computed the respective number of degree days from 1 March (using 40^{o}F as the threshold level) for the two sites. Based on the similarity of number of degree days on the two sites on both 18 April and 6 May, we accepted the two sites as being phenologically analogous on the two dates at the two sites.

One test of the success of extension of wheat yield relations involved calculating a mean square error (MSE) for a regression between Landsat data and yield on a particular site and comparing that with a MSE calculated when yield is predicted on that same site using a relationship developed on another site. These two values of MSE furnished the basis for an "F-statistic", the ratio of the MSE for the extended relation to the MSE for the base equation. The larger the F-ratio, the worse the prediction of individual field yields was compared to the base prediction of yield.

Another statistical test performed was to determine how well the average yield for all fields was predicted. This test, a "t-test" was then computed as:

$$t = \frac{\bar{\hat{Y}} - \bar{Y}}{s/\sqrt{n}}$$

where

$\bar{Y}$ = average value of yield

$\bar{\hat{Y}}$ = average predicted value of yield

$$s^2 = \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2/n-1$$

The null hypothesis is $\bar{\hat{Y}} - \bar{Y} = 0$, or that the mean values of actual and Landsat-predicted yield are the same. The larger the t-value, the less likely is the hypothesis to be true.

F and t tests were computed for data that was not normalized in any way, in order to determine the severity of the problem of using non-normalized Landsat data. F and t tests were subsequently computed for the normalization techniques mentioned previously, namely SQ75 and TVI. The results are presented in Table 8.

While there is much about these results that remains to be analyzed, a few aspects which have been examined should be mentioned. The two

TABLE 8. SUMMARY OF YIELD PREDICTION EXTENSIONS
1976 FINNEY AND ELLIS SITES

Yield Extension From	Yield Extension To	Landsat Predictor	Bias (Bushels) $(\hat{Y}_{Pred} - \hat{Y}_{True})$	RMS Error (Bushels)	t-test	F-test
Finney 6 May	Ellis 6 May	4 Bands	-4.02	7.70	3.53*	1.79
		SQ75	1.60	7.00	1.48	1.37
		TVI	0.91	5.97	0.99	1.06
Ellis 6 May	Finney 6 May	4 Bands	0.28	8.12	0.28	1.12
		SQ75	2.04	8.79	1.85	1.41
		TVI	0.55	7.76	0.55	1.05
Finney 18 April	Ellis 18 April	4 Bands	-1.74	10.28	1.38	3.01*
		SQ75	6.77	12.14	3.62*	3.95*
		TVI	4.81	9.03	3.46*	2.32*
Ellis 18 April	Finney 18 April	4 Bands	0.23	9.10	0.20	1.51
		SQ75	2.15	10.18	1.64	1.57
		TVI	1.17	9.29	0.98	1.33

* Statistically significantly different at 5% level.

data normalizations (SQ75, TVI) perform better than the non-normalized bands for the extension from 6 May Finney to 6 May Ellis. However, depending on the test used for the other three extensions there is no clear superiority of using a data transformation (TVI, SQ75) rather than using all four bands. For these data it thus appears that differences in external effects, such as atmospheric conditions, soil reflectance, sun angle, etc., were not sufficiently serious that the benefit of using a transformation to reduce such effects exceeded the concurrent penalty due to using less yield-predictive information than was available.

Extensions of yield prediction were more successful when 6 May data was used than when 18 April data was used. In addition, there is at this time an unexplained tendency for the extensions from Ellis to Finney to be more successful than extensions from Finney to Ellis.

The results indicate that for these tests, no matter which transform is better correlated with yield locally, TVI tends to perform better in the yield extensions than SQ75. This situation has not always been found previously (e.g., August-November 1976 Quarterly Report).

Little consistency in results has been achieved thus far in analysis of performance of yield prediction extension. The reasons for the discrepancies are not always clear. It may be that procedures that are generally optimum can be discovered only be development of a larger base of tests of candidate procedures.

Considering the implications of these results in terms of a large area survey, however, the picture is not discouraging. It seems likely that the RMS error we obtained on the 6 May yield prediction extension data is similar to what one may achieve in a large area survey situation. The average RMS error over all 6 May extensions (all three techniques) was 7.6 bushels/acre, on a field-by-field basis. Because of the statistical central limit theorem, the error in an estimate of average yield (bias) based on N fields would be smaller by the factor $1/\sqrt{N}$ if error in predicted yield is normally distributed around zero. The actual average absolute value of bias in the six extensions (N = 40 for Ellis, N = 58 for Finney) is only 1.6 bushels per acre.

8.0 DIRECT WHEAT PRODUCTION FORECASTS

In order to bring the wheat yield estimation technology addressed by this study into a more directly applicable framework, we are examining the possibility of using the technology in formulating and testing methods to directly forecast total wheat production (in bushels) on a regional basis. Our work thus far has suggested a new and simple approach to production forecasting which may overcome certain troublesome problems in some of the existing approaches. The existing approaches tend to separate the task of forecasting into two separate subsystems consisting of: (1) wheat acreage determination; and (2) regional average determination of per acre yield. The approach discussed below could make it possible to determine acreage and yield simultaneously on a pixel-by-pixel basis, using early-season Landsat data, with a single processing step. Thus it may become possible to survey large areas, such as a state or country, much more economically than at present, and achieve more timely information. What follows is a discussion of the rationale of the suggested approach, and a demonstration of its initial implementation.

The basic idea in the direct wheat production approach using Landsat data is that an appropriate value of yield (per unit area) can be determined for each pixel in the scene, without the need to specify that the pixel is wheat, and that production can be determined as

$$\text{Production} = \sum_{i=1}^{n} \text{yield}_i \times (\text{area of a pixel})$$

where i numbers the set of n pixels covering the area of interest.

We have previously shown that several Landsat transforms are good indicators of green vegetative cover, and that cover, as so measured, in turn is strongly related to wheat yield. An additional fact, which is further discussed below is that in winter wheat regions such as Kansas, wheat tends to develop significant green cover sooner than most non-wheat fields. Thus, if a yield-predictive relation (developed on wheat fields) is applied to non-wheat pixels, a very low yield indication would be expected, and might be a negligible source of error. If applied to pixels falling on a boundary between wheat and non-wheat, an appropriate intermediate value of green cover and thus weighted average yield would be estimated. This intermediate value of yield estimates times the area per pixel could approximate the total amount of wheat production represented by the pixel, which covers an area only partially planted to wheat. Thus, in all cases pixels tend to contribute only their fair share of the total production estimate.

As mentioned above, our approach depends on the hypothesis that non-wheat fields tend to have a smaller measure of green vegetative cover than wheat fields. Non-wheat classes should be largely separable from wheat using a Landsat indicator of green vegetative cover. In order to test these hypothesis, we selected as a green measure SQ75. The measure SQ75 was computed for all sufficiently large fields in the Finney County, Kansas site using 6 May 1976 Landsat data. A threshold was selected to optimally distinguish wheat from non-wheat using SQ75. As a result, four of 58 wheat fields fell below the threshold, and two of 38 non-wheat fields fell above, giving a classification accuracy of 93.8% correct. A comparison of wheat and non-wheat histograms illustrating the separability is given in Figure 8. The same procedure applied to 6 May 1976 Landsat data for the Ellis County site resulted in an overall classification accuracy of 91.9%. Similar indications of the utility of Landsat green cover measures for wheat recognition have been demonstrated at ERIM [6]. We therefore assume that an early-season green cover measure can give a reasonably accurate classification of wheat in some winter wheat regions.

FIGURE 8. SEPARABILITY OF WHEAT FROM NON-WHEAT USING HISTOGRAMS OF THE SQ75 TRANSFORMATION. FINNEY SITE, 6 MAY 1976.
(Each x = 1 Field)

SQ75	Non-Wheat	Wheat
.66000	0 +	0 +
.68000	3 +XXX	0 +
.70000	15 +XXXXXXXXXXXXXXX	0 +
.72000	11 +XXXXXXXXXXX	2 +XX
.74000	4 +XXXX	2 +XX
.76000	2 +XX	0 +
.78000	2 +XX	3 +XXX
.80000	0 +	1 +X
.82000	0 +	0 +
.84000	1 +X	5 +XXXXX
.86000	0 +	0 +
.88000	0 +	3 +XXX
.90000	0 +	8 +XXXXXXXX
.92000	0 +	4 +XXXX
.94000	0 +	7 +XXXXXXX
.96000	0 +	3 +XXX
.98000	0 +	2 +XX
1.0000	0 +	4 +XXXX
1.0200	0 +	1 +X
1.0400	0 +	1 +X
1.0600	0 +	2 +XX
1.0800	0 +	3 +XXX
1.1000	0 +	3 +XXX
1.1200	0 +	0 +
1.1400	0 +	1 +X
1.1600	0 +	0 +
1.1800	0 +	0 +
1.2000	0 +	2 +XX
1.2200	0 +	0 +
1.2400	0 +	1 +X
1.2600	0 +	0 +

Next we examined a simple method of direct production estimation. Again using SQ75 as a green cover measure, we obtained a yield predictive relation based on the wheat fields in a 4 x 6 mile training area chosen within the Finney site using 6 May 1976 Landsat data. Using the relation, we computed an estimate of yield for each pixel in a test region consisting of the remaining 1 x 6 mile area in the site. The yield from each pixel times the acreage associated with a pixel was summed over all pixels in the test region, giving the total production estimated for the 1 x 6 mile test segment. In doing so, it had been assumed that yield attributed to non-wheat pixels may be negligible, although the assumptions had not yet been checked.

As a result, the production estimate for the test area was 53,900 bushels, compared to the "true" production (as computed from farmer reported production information) of 40,600 bushels, a 33% overestimate. On examining the assumption of negligible production from non-wheat fields, we found that the average yield/acre associated with non-wheat fields was about 5 bushels per acre. Although this is a rather small yield (compared to typical yields of 30-40 bu/acre and maximum yields around 60 bu/acre), it is multiplied by a very large number of pixels (acres), and so leads to an overestimate of production on the order of what was observed.

Due to the above consideration, we modified the technique to account for the production improperly associated with non-wheat, by selecting a threshold below which a pixel is assumed to be non-wheat or wheat sufficiently marginal as to be possibly not worth harvesting. Initially we chose a threshold so as to approximately make compensating errors in acreage estimation. More specifically, a threshold was determined so as to minimize the difference between the number of wheat pixels below the threshold and the number of non-wheat pixels above the threshold in the training region.

When production estimates were made as described previously, but using a threshold determined using the fields in the training area of the Finney site, we obtained a production estimate of 42,700 bushels, compared with the actual 40,600 bushels, which represents an error of only 5.2%. In addition, we applied the same procedure to the same site using 18 April 1976 Landsat data, and to a different site (Ellis County, Kansas) using 6 May 1976 Landsat data. For the Ellis site a 6 square mile training area and a separate 3 square mile test area were used. The resulting production estimate are shown in Table 9.

TABLE 9. RESULTS FROM SIMPLE DIRECT WHEAT PRODUCTION ESTIMATION PROCEDURE

Site	Landsat Overpass	True Production (10^3 Bushels)	ERIM Estimate (10^3 Bushels)	Error (%)
Finney	6 May 76	40.6	42.7	5.2%
Finney	18 Apr 76	40.6	42.8	5.4%
Ellis	6 May 76	27.9	24.7	11.5%

Preliminary indications based on the three test results give encouragement that the direct wheat production approach using early-season Landsat data might produce reasonable results. Many more tests in different situations will have to be performed in order to assess the consistency in performance. It is anticipated that variations in desired approach or acceptable calibration may occur in other situations, and that stratification of data may be required.

However, the approach does address some problems that may exist in present methods. As indicated in Section 6, local variations in yield can possibly be accounted for with greater precision using Landsat data than using meteorological data. The difficulty in locating field boundaries on Landsat data for determination of wheat acreage is alleviated since all pixels are included in the proposed new technique. Small or irregularly shaped fields can contribute to the acreage and production estimate even if not a single pixel falls completely within the field boundary. Furthermore, large bare areas within wheat fields will be assigned little or no yield, thereby giving approximately the correct production, without a decision having to be made as to whether the area should be assigned to wheat acreage or not. Finally, marginal wheat fields, ones which are not likely to be harvested, will not be included in early season production forecasts if they fall below the green indicator threshold.

There are some indications that these potential desirable features of the direct wheat production approach are being fulfilled. For example, there were several wheat fields in our Finney test for which no "pure" pixels could be obtained. That is, all pixels covering these fields were on the field boundary, or very nearly so. One such field had a farmer reported production of 1001 bushels and an area of 32.7 acres. Even though not a single pure pixel was present, production of 732 bushels was estimated for this field, based just on the pixels whose centers fell within the field boundaries.

In the Ellis site there was a wheat field which was not harvested because the stand was too sparse. Every pixel within that field boundary had a green transform value less than the minimum threshold. Therefore, even though the field was wheat it could not have contributed to a production estimate, which is the desired result in this case since no wheat was produced on this field.

9.0 FUTURE PLANS

During the next reporting period we plan to extend our activities to demonstrate Landsat-based yield estimation over an entire Crop Reporting District (CRD). We intend to process Landsat data from selected areas in that district, estimate yield for each area and aggregate the results to obtain an average yield estimate for the entire CRD. This Landsat-derived yield estimate will then be compared with the available Crop and Livestock Reporting Service estimate, and also with an estimate based on an agrometeorological yield model. We also plan to further examine our direct production estimation technique.

REFERENCES

1. Colwell, J., "Bidirectional Spectral Reflectance of Grass Canopies for Determination of Above Ground Standing Biomass", Ph.'D. Dissertation, University of Michigan, Ann Arbor, Michigan, 1973.

2. Nalepka, R., J. Colwell, D. Rice, "Wheat Productivity Estimates Using Landsat Data", Type II Progress Report #114800-25-L on NASA Contract NAS5-22389, 1976.

3. Malila, W., R. Cicone, and J. Gleason, "Wheat Signature Modeling and Analysis for Improved Training Statistics", ERIM Report Number 109600-66-F, NASA Contract Number NAS9-14123, May 1976.

4. Nalepka, R., J. Colwell, D. Rice, "Wheat Productivity Estimates Using Landsat Data", Type II Progress Report #114800-12-L on NASA Contract NAS5-22389, 1976.

5. Feyerherm, A., "Prediction of Wheat Yield", Quarterly Progress Report, NASA Contract NAS9-14533, August 1976.

6. Nalepka, R., et al, "Analysis of Scanner Data for Crop Inventory", Task 3 ERIM Report 122700-23-L on NASA Contract NAS9-14988, March 1976.